

# Small-noise analysis and symmetrization of implicit Monte Carlo samplers

Jonathan Goodman<sup>\*</sup>

Kevin K. Lin<sup>†</sup>

Matthias Morzfeld<sup>‡</sup>

October 21, 2014

## Abstract

Implicit samplers are algorithms for producing independent, weighted samples from multi-variate probability distributions. These are often applied in Bayesian data assimilation algorithms. We use Laplace asymptotic expansions to analyze two implicit samplers in the small noise regime. Our analysis suggests a symmetrization of the algorithms that leads to improved (implicit) sampling schemes at a relatively small additional cost. Computational experiments confirm the theory and show that symmetrization is effective for small noise sampling problems.

## 1 Introduction

Markov chain Monte Carlo (MCMC) techniques are widely used for sampling complicated distributions. However, some data assimilation methods rely on independent samples from known distributions [5, 10, 12, 26].

---

<sup>\*</sup>Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

<sup>†</sup>School of Mathematics, University of Arizona, Tucson, AZ 85721, USA

<sup>‡</sup>Department of Mathematics, University of California at Berkeley and Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Weighted direct samplers give independent samples from a proposal distribution that is not the target distribution, and compensate for this with a random weight factor (see e.g. [6] and references there). The variance of the weight factor determines the quality of the sampler [2, 10, 18].

This paper studies two weighted direct samplers.

One, the *linear map* method, has been proposed independently several times and has several names in the literature; see, e.g., [8], and also [1] for a similar method. The other is the *random map* method, which was proposed in [23]. The linear and random map methods can both be viewed as examples of *implicit samplers* [3, 8, 9]. We introduce a small noise parameter,  $\varepsilon$ , similar to that of [27, 28], and analyze the performance of these algorithms in sampling general smooth probability densities on finite-dimensional spaces in the limit  $\varepsilon \rightarrow 0$ . The methods we study use a Gaussian approximation to the target distribution, which is valid in the small noise limit. Many data assimilation applications are in the small noise regime.

We study a standard quality measure of weighted direct samplers. Our analysis consists in calculating *error constants*, which are the coefficients of the leading powers of  $\varepsilon$  in the small noise asymptotic expansion of the quality measures. As long as simple smoothness hypotheses are satisfied, the error constants for the linear and random map methods differ by a factor that depends only on the dimension. This factor converges to one as the dimension goes to infinity.

The form of the error constant suggests that a symmetrization may remove the leading error term. We study symmetrized versions of the linear and random map methods to confirm this. The error is one order smaller in  $\varepsilon$ . The error constants are not exactly proportional, but their ratio does converge to one as the dimension converges to infinity. We present computational experiments that confirm the small noise asymptotic calculations. The numerical experiments further demonstrate that the symmetrized methods are more accurate in the small noise regime, and show that the symmetrized methods may perform significantly better than the corresponding “simple” methods even when the noise is not so small.

This paper is organized as follows: in Section 2, we set up the notation, present the algorithms, explain how they can be symmetrized, and summarize our theoretical results. Section 3 describes two general technical tricks that simplify the asymptotic analysis. The relatively simple deriva-

tions of the linear map results are in Section 4. These depend on Subsection 3.1 only. The ideas in Subsection 4.3 are needed only for the explicit error constant formulas for the first computational example in Section 6. The analysis in Section 5 of random map methods also uses the formula derived in Subsection 3.2. Section 6 describes numerical experiments on two test problems which confirm the asymptotic theory in detail. It may be read without the theoretical Sections 3, 4, and 5. Section 7 summarizes our views of these results and puts them in context.

## 2 Algorithms, symmetrization and main results

In this section, we describe the sampling algorithms to be studied in the rest of the paper. We introduce the small-noise scaling used in the analysis in Section 2.3, where we also state our main theoretical scaling results.

The following notation is used throughout the paper. Let  $f(x)$  and  $r(x)$  be two functions on  $\mathbb{R}^d$ . We write  $f \propto r$  if  $f(x) = Cr(x)$  for some fixed  $C$ . If distributions depend on  $\varepsilon$ , we write  $f(x, \varepsilon) \propto r(x, \varepsilon)$  if there is a  $C_\varepsilon$  with  $f(x, \varepsilon) = C_\varepsilon r(x, \varepsilon)$ . For any non-negative  $f$  with  $0 < \int f(x) dx < \infty$ , there is a probability density  $p \propto f$ . We write  $X \sim p$  if  $X$  is a random variable whose probability density is  $p$ . We say a random variable  $X \sim q$  together with a non-negative weight function  $w$  is a *weighted sample* of a given probability density  $p$  if

$$E_p(u(X)) = \frac{E_q(u(X) \cdot w(X))}{E_q(w(X))}, \quad (1)$$

for every bounded continuous function  $u$ . A *weighted sampler* of  $p$  with *proposal*  $q$  is a stochastic algorithm that produces  $X \sim q \propto g$ . It is a *direct sampler* if successive samples are independent. The direct samplers we consider have deterministic weight functions

$$w(x) = \frac{f(x)}{g(x)} \propto \frac{p(x)}{q(x)}. \quad (2)$$

We assume  $f$  and  $g$  may be evaluated, but the normalizing constants may be unknown (as is typically the case in applications).

A perfect sampler would have a constant weight function  $w = C$ , which would force  $p = q$ . We measure the quality of a weighted sampler

by the non-dimensionalized deviation of  $w$  from a constant:

$$R := \frac{E_q(w(X)^2)}{E_q(w(X))^2}, \quad Q := R - 1. \quad (3)$$

The quality measure  $Q$  was also used in [28], and several motivations for it are given in [2,4,10,19]. In particular, a heuristic relates a collection of  $N$  independent weighted samples to  $N/R$  independent un-weighted samples, making  $N/R$  an effective sample size. A small  $Q$  is important in recursive particle filter algorithms. There, probability densities are sampled recursively, as the data are collected, and the weights accumulate as a product of the weights at each step. Thus, the  $R$  of the product can grow rapidly if the  $Q$  of each of the factors is not small. Both algorithms analyzed here have the property that  $Q \rightarrow 0$  as  $\varepsilon \rightarrow 0$ ; the question that concerns us is the rate of convergence.

The methods we consider sample

$$p(x) \propto f(x) = e^{-F(x)}. \quad (4)$$

We assume that  $F$  is smooth and has a single global minimum, which is non-degenerate. Unless otherwise stated, we also assume (without loss of generality) that the minimum is located at  $x = 0$ , and that  $F(0) = 0$ . We write a Taylor expansion of  $F$  near zero as

$$F(x) = \frac{1}{2}x^t H x + C_3(x) + \cdots + C_6(x) + O(|x|^7), \quad (5)$$

where  $H$  is the Hessian matrix of  $F$  at  $x = 0$ , and  $C_k(x)$  is the homogeneous polynomial of degree  $k$

$$C_k(x) = \frac{1}{k!} \sum_{|\alpha|=k} x^\alpha \partial_x^\alpha F(0). \quad (6)$$

The random map samplers also require that certain equations related to  $F$  have unique and well behaved solutions (see below).

## 2.1 Simple and symmetrized linear map methods

The simple linear map method uses  $X \sim \pi$ , where  $\pi$  is the local Gaussian approximation that uses the first term on the right of (5),

$$\pi(x) \propto e^{-x^t H x / 2}. \quad (7)$$

Direct Gaussian sampling algorithms make this possible. Using (2) with  $f = e^{-F}$  and  $g = e^{-x^t H x / 2}$ , we find the weight function

$$w(x) = e^{-F(x) + x^t H x / 2} . \quad (8)$$

The simple linear map Monte Carlo algorithm to estimate  $E_p(u(X))$  is:

1. Generate  $N$  independent Gaussian samples  $X_k \sim \pi$
2. Compute weights  $W_k = w(X_k)$  using (8)
3. Compute the estimator

$$\frac{\sum_{k=1}^N u(X_k) w(X_k)}{\sum_{k=1}^N w(X_k)} ,$$

of  $E_p(u(X))$  .

In practice, the minimizer of  $F$  will not be at 0 , nor will its Hessian at the minimum be the identity. It will be necessary to first find  $x_* = \operatorname{argmin} F(x)$  and evaluate  $H(x_*)$ , the Hessian of  $F$  at the minimum. This can be a time-consuming step.

As we will see below, the leading-order term in the  $\varepsilon$ -expansion of  $Q$  depends only on  $C_3$  (see equation (36)). This is not surprising, as the simple method is based on the approximation  $F(x) \approx \frac{1}{2} x^t H x$ , and  $C_3(x)$  is the largest correction. Since  $C_3$  is an odd function of  $x$ , one may hope that the leading error term can be removed by a symmetrization related to the classical Monte Carlo trick of antithetic variates [15,16]. Here we present a symmetrized linear map method; we will verify in Section 4.2 that it removes the principal error term in the small noise limit.

The symmetrized linear map method is as follows: first draw  $\xi \sim \pi$  as before, and evaluate the linear map weights (8) for  $\xi$  and for  $-\xi$ . Note that  $\pi(\xi) = \pi(-\xi)$  so these weights are

$$w_+ = \frac{f(\xi)}{\pi(\xi)} , \quad w_- = \frac{f(-\xi)}{\pi(\xi)} . \quad (9)$$

Then return  $X = \xi$  or  $X = -\xi$  with probabilities

$$p_+ = \frac{w_+}{w_+ + w_-} , \quad p_- = \frac{w_-}{w_+ + w_-} . \quad (10)$$

These probabilities have a particle filter interpretation. Consider  $(\xi, w_+)$  and  $(-\xi, w_-)$  to form a two element weighted ensemble. The formulas (10) are the probabilities that would be used to sub-sample this to a one element un-weighted ensemble [2].

To find the weight function of the symmetrized linear map method we must identify  $q_s(x)$ , the probability density of  $X$ . There are two ways to generate  $X = x$  (using the convention that  $x$  is a possible value of the random variable  $X$ ). One way is to propose  $\xi = x$  and then take the  $+$  choice in (10). The other way is to propose  $\xi = -x$  and then take the  $-$  choice. The probability density for  $\xi = x$  is  $\pi(x)$ . The density for  $-\xi$  is  $\pi(-\xi) = \pi(\xi)$ . The probability to get  $x$  if  $x$  was proposed is

$$p_+(x) = \frac{w(x)}{w(x) + w(-x)} . \quad (11)$$

The probability to get  $x$  if  $-x$  was proposed is the same, since

$$\begin{aligned} p_-(-x) &= \frac{w_-(-x)}{w_+(-x) + w_-(-x)}, \\ &= \frac{w(-(-x))}{w(-x) + w(-(-x))}, \\ &= p_+(x) . \end{aligned} \quad (12)$$

Therefore, the pdf of  $X$  is

$$\begin{aligned} q_s(x) &= \pi(x)p_+(x) + \pi(-x)p_-(-x), \\ &= \pi(x) \frac{2w(x)}{w(x) + w(-x)} . \end{aligned} \quad (13)$$

Moreover, if  $\pi(x)$  is a normalized probability density, then  $q_s$  is also normalized. This can be seen by using the right side of (13)

$$\begin{aligned} \int q_s(x) dx &= \frac{1}{2} \int (q_s(x) + q_s(-x)) dx, \\ &= \frac{1}{2} \int \frac{2\pi(x)}{w(x) + w(-x)} (w(x) + w(-x)) dx, \\ &= 1 . \end{aligned} \quad (14)$$

The weight function (2) for the symmetrized method is thus

$$w_s(x) \propto \frac{p(x)}{q_s(x)} \propto \frac{\pi(x)w(x)}{\pi(x)\frac{2w(x)}{w(x)+w(-x)}} = \frac{w(x) + w(-x)}{2} . \quad (15)$$

The simple linear map sampler and the symmetrized sampler have different symmetries. The simple sampler has a symmetric proposal density and non-symmetric weight. The symmetrized sampler has a non-symmetric proposal density,  $q_s(x) \neq q_s(-x)$ , but a symmetric weight function. Intuitively one can therefore expect that the quality measure of the symmetrized method is better because a symmetric weight function is “more nearly constant” for small  $x$ , particularly in the small noise regime described below.

## 2.2 Simple and symmetrized random map methods

The *simple* (as opposed to *symmetrized*) random map method is described in [23]. We review the method here for notation and completeness.

One first samples  $\xi \sim \pi$ , and then chooses

$$X = \lambda(\xi)\xi . \quad (16)$$

The stretch factor  $\lambda(\xi) \geq 0$  is defined implicitly via

$$F(\lambda(\xi)\xi) = \frac{1}{2}\xi^t H \xi . \quad (17)$$

The random map algorithm gets its name from the map  $\xi \mapsto X$ . To ensure the correctness of the algorithm, we need to assume that equation (17) has a unique solution  $\lambda > 0$  for every  $\xi \neq 0$ . This will be the case if, e.g., every level set (except the zero level set) of  $F$  is “star-shaped,” i.e., for every  $c > 0$ , every straight line through 0 intersects the level set  $F^{-1}(c)$  transversely at exactly two points.

To determine the weight function of the random map method, note that if  $\xi \sim \pi$  and  $X = x(\xi)$ , then  $X$  has probability density

$$q(x(\xi)) = \pi(\xi) \left| \det \left( \frac{\partial \xi}{\partial x} \right) \right| , \quad (18)$$

so that we find the weight  $w$  from (2) to be

$$w(\xi) = \left| \det \left( \frac{\partial x}{\partial \xi} \right) \right| , \quad (19)$$

choosing the arbitrary implicit constant to be equal to 1 here.

The Jacobian determinant is

$$w(\xi) = \lambda(\xi)^{d-1} \frac{\xi^t H \xi}{|\xi^t \nabla_x F(\lambda(\xi) \xi)|} . \quad (20)$$

To see this, note that the Jacobian matrix is obtained by differentiating (16)

$$\frac{\partial x}{\partial \xi} = \xi [\nabla \lambda(\xi)]^t + \lambda(\xi) I, \quad (21)$$

where  $\nabla \lambda$  is the column vector with entries  $\partial_{\xi_j} \lambda(\xi)$ . The determinant identity

$$\det(\lambda I + A) = \lambda^d + \lambda^{d-1} \text{tr}(A) + \dots , \quad (22)$$

gives

$$\det\left(\frac{\partial x}{\partial \xi}\right) = \lambda^d + \lambda^{d-1} \xi^t \nabla \lambda , \quad (23)$$

where the terms of order  $\lambda^{d-2}$  and lower vanish because  $\xi [\nabla \lambda(\xi)]^t$  is a matrix of rank one. A calculation (given just below) gives

$$\xi^t \nabla_\xi \lambda = \lambda \left( \frac{\xi^t H \xi}{x^t \nabla_x F} - 1 \right) , \quad (24)$$

which immediately leads to (20). To verify (24), we differentiate (17) with respect to  $\xi_i$ :

$$\sum_j \partial_{x_j} F(\lambda(\xi) \xi) \partial_{\xi_i} [\lambda(\xi) \xi_j] = (H \xi)_i , \quad (25)$$

$$\sum_j \partial_{x_j} F(\lambda(\xi) \xi) \left( \frac{\partial \lambda(\xi)}{\partial \xi_i} \xi_j + \lambda(\xi) \delta_{ij} \right) = (H \xi)_i . \quad (26)$$

We multiply by  $\xi_i$ , sum over  $i$ , and use the relations  $\lambda(\xi) \xi = x$ , and  $\xi = \frac{1}{\lambda(\xi)} x$ :

$$\xi^t \nabla_x F(\lambda(\xi) \xi) \xi^t \nabla_\xi \lambda(\xi) + \lambda(\xi) \xi^t \nabla_x F(\lambda(\xi) \xi) = \xi^t H \xi, \quad (27)$$

$$\frac{1}{\lambda(\xi)} x^t \nabla_x F(\lambda(\xi) \xi) \xi^t \nabla_\xi \lambda(\xi) + x^t \nabla_x F(\lambda(\xi) \xi) = \xi^t H \xi . \quad (28)$$

Solving for  $\xi^t \nabla_x \lambda(\xi)$  gives (24).

Our symmetrization of the simple random map method is a natural adaptation of the symmetrization of the simple linear map method. There are three steps:



1. Generate a sample  $\xi \sim \pi$ .
2. Compute  $x_+ = \lambda(\xi) \cdot \xi$  and  $x_- = \lambda(-\xi) \cdot (-\xi)$ , each using (17).
3. Use  $x = x_+$  with probability  $p_+(\xi) := w(\xi)/(w(\xi) + w(-\xi))$ . Otherwise use  $x_-$ .

The arguments leading to (13) and (15) apply here too. The probability density of  $X$  produced by the symmetrized random map method is thus

$$q_s(x) = \frac{2}{w(\xi(x)) + w(-\xi(x))} e^{-F(x)}, \quad (29)$$

where  $w(\xi)$  is the weight of the simple random map method. The weight function for the symmetrized random map method is

$$w_s(\xi) = \frac{w(\xi) + w(-\xi)}{2}, \quad (30)$$

where  $w$  is the weight of the simple method (20).

## 2.3 Summary of small noise theory

The small noise problem concerns the scaled density

$$p(x) \propto f(x) = e^{-F(x)/\varepsilon}. \quad (31)$$

Recursive particle filter applications often call for proposal distributions roughly of the form (31). When the noise parameter  $\varepsilon$  is small, most of the probability in  $p$  is near the point of maximum probability, which we continue to take to be  $x_* = 0$ . Therefore  $F(x) \approx \frac{1}{2}x^t H x$  (see (5)) may be a useful approximation.

We state and derive the small noise theory using a standard scaling,

$$\tilde{x} = \varepsilon^{1/2} x. \quad (32)$$

This scales the terms in the Taylor expansion (5) as

$$\frac{F(\tilde{x})}{\varepsilon} = \frac{1}{2}\tilde{x}^t H \tilde{x} + \varepsilon^{1/2} C_3(\tilde{x}) + \varepsilon C_4(\tilde{x}) + \varepsilon^{3/2} C_5(\tilde{x}) + \varepsilon^2 C_6(\tilde{x}) + O(\varepsilon^{5/2}). \quad (33)$$

The target density therefore satisfies

$$p(\tilde{x}) \propto \exp \left( -\tilde{x}^t H \tilde{x} / 2 - \varepsilon^{1/2} C_3(\tilde{x}) - \varepsilon C_4(\tilde{x}) - O(\varepsilon^{3/2}) \right). \quad (34)$$

For the rest of the theory, we assume  $p$  satisfies (34). Following common practice, we drop the tilde.

Our theoretical results take the form of asymptotic approximations of  $Q$  defined in (3). The simple linear map and random map methods have the scaling

$$Q = \varepsilon A + O(\varepsilon^{3/2}). \quad (35)$$

The error constants are

$$A = E_\pi(C_3(X)^2) \quad (\text{simple linear map}), \quad (36)$$

$$A = \frac{(1+d)^2}{(2+d)(4+d)} E_\pi(C_3(X)^2) \quad (\text{simple random map}). \quad (37)$$

The error scaling for the symmetrized methods is

$$Q = \varepsilon^2 B + O(\varepsilon^{5/2}), \quad (38)$$

with error constants of the form

$$B = \text{var}_\pi \left( C_4 - \frac{1}{2} C_3^2 \right) \quad (\text{symmetrized linear map}), \quad (39)$$

$$B = \text{var}_\pi \left( C_4 - \frac{1}{2} C_3^2 \right) + c_d \cdot K \quad (\text{symmetrized random map}). \quad (40)$$

Here  $c_d = O(1/d)$ , and  $K$  is a possibly dimension-dependent constant depending on  $F$ . The exact form is given in Section 5.2.

We have the following conclusions. For both methods,  $Q \rightarrow 0$  in the small noise limit  $\varepsilon \rightarrow 0$ . This is perhaps not surprising because the Gaussian approximation  $\pi$  becomes exact in this limit. On the other hand, this property cannot be taken for granted in general; see, e.g., [27].

The simple linear and random map methods have the same order as  $\varepsilon \rightarrow 0$ , and the symmetrized methods have a higher order. Thus, for any fixed problem and for sufficiently small  $\varepsilon$ , the error constants of the symmetrized methods are significantly smaller than the error constants for the corresponding simpler methods. The ratio of the error constants for the linear and random map methods depends only on the dimension. This factor converges to 1 as  $d \rightarrow \infty$ . Thus, in the limits  $\varepsilon \rightarrow 0$  and  $d \rightarrow \infty$ , the random map methods lose their advantages over the linear map methods.

### 3 Analysis tools

Here we describe two tools that we will use in the analysis of the linear and random map methods.

#### 3.1 The variance lemma

The variance lemma is a simple way to understand some cancellations that occur in computing  $Q$  for small  $\varepsilon$ . It applies to functions  $u(x, \varepsilon)$  of the form

$$u(x, \varepsilon) = 1 + \varepsilon^r u_1(x) + \varepsilon^{2r} u_2(x) + O(\varepsilon^{3r}) .$$

It states that if

$$Q = \frac{E(u(X, \varepsilon)^2)}{E(u(X, \varepsilon))^2} - 1 , \quad (41)$$

then

$$Q = \varepsilon^{2r} \text{var}(u_1(X)) + O(\varepsilon^{3r}) . \quad (42)$$

The variance formula (42) does not depend on the distribution of  $X$ , except that the same distribution must be used throughout. Expectations of  $u_2$  appear at  $O(\varepsilon^{2r})$  in the numerator and denominator of (41) but they cancel in the ratio to leading order.

The verification is straightforward. The numerator in (42) is

$$\begin{aligned} E(u(X, \varepsilon)^2) &= E(1 + 2\varepsilon^r u_1 + \varepsilon^{2r} (u_1^2 + 2u_2) + O(\varepsilon^{3r}), \\ &= 1 + 2\varepsilon^r E(u_1) + \varepsilon^{2r} (E(u_1^2) + 2E(u_2)) + O(\varepsilon^{3r}) . \end{aligned}$$

The denominator is

$$\begin{aligned} E(u(X, \varepsilon))^2 &= (1 + \varepsilon^r E(u_1) + \varepsilon^{2r} E(u_2) + o(\varepsilon^{2r}))^2 , \\ &= 1 + 2\varepsilon^r E(u_1) + \varepsilon^{2r} (E(u_1)^2 + 2E(u_2)) + O(\varepsilon^{3r}) . \end{aligned}$$

Therefore,

$$\begin{aligned} Q &= \frac{1 + 2\varepsilon^r E(u_1) + \varepsilon^{2r} (E(u_1^2) + 2E(u_2)) + O(\varepsilon^{3r})}{1 + 2\varepsilon^r E(u_1) + \varepsilon^{2r} (E(u_1)^2 + 2E(u_2)) + O(\varepsilon^{3r})} - 1 \\ &= [1 + 2\varepsilon^r E(u_1) + \varepsilon^{2r} (E(u_1^2) + 2E(u_2)) + O(\varepsilon^{3r})] \\ &\quad \times [1 - 2\varepsilon^r E(u_1) + \varepsilon^{2r} (3E(u_1)^2 - 2E(u_2)) + O(\varepsilon^{3r})] - 1 \\ &= \varepsilon^{2r} (E(u_1^2) - E(u_1)^2) + O(\varepsilon^{3r}) . \end{aligned}$$

Note that this conclusion depends on the existence of a function  $u_2(x)$ , but it does not depend on what  $u_2$  is.

### 3.2 Evaluating rational Gaussian expectations

The random map analysis in Section 5 leads to Gaussian expectations of the form

$$E_\pi \left( \frac{C(\xi)}{|\xi|^{2r}} \right) ,$$

where  $C$  is a homogeneous polynomial of some degree. These are related to expectations of  $C$ . In fact, if  $f(\xi)$  is homogeneous of degree  $q$ , then

$$(q + d)E_\pi(f(\xi)) = E_\pi(|\xi|^2 f(\xi)) . \quad (43)$$

Taking  $C$  of degree  $p$ , and  $f(\xi) = C(\xi)/|\xi|^2$  or  $f = C(\xi)/|\xi|^4$ , gives  $q = p - 2$  or  $q = p - 4$ , and

$$E_\pi \left( \frac{C(\xi)}{|\xi|^2} \right) = \frac{1}{p - 2 + d} E_\pi(C(\xi)) , \quad (44)$$

or (iterating twice)

$$E_\pi \left( \frac{C(\xi)}{|\xi|^4} \right) = \frac{1}{(p - 4 + d)(p - 2 + d)} E_\pi(C(\xi)) . \quad (45)$$

This result, which may be derived as a  $\Gamma$  function identity, is surely not new.

We give an elementary derivation that uses the function

$$I(r) = \int f(r\xi) e^{-|\xi|^2/2} d\xi .$$

On one hand,

$$I(r) = r^q \int f(\xi) e^{-|\xi|^2/2} d\xi = r^q I(1) .$$

On the other hand, we can change variables with  $r\xi = \eta$  to get

$$I(r) = \int f(\eta) e^{-|\eta|^2/(2r^2)} \frac{1}{r^d} d\eta .$$

Now differentiate with respect to  $r$  and set  $r = 1$ :

$$\begin{aligned}
qr^{q-1}I(1) &= I'(r) \\
&= \frac{1}{r^3} \int f(\eta)|\eta|^2 e^{-|\eta|^2/(2r^2)} \frac{d\eta}{r^d}, \\
&\quad - d \int f(\eta)|\eta|^2 e^{-|\eta|^2/(2r^2)} \frac{d\eta}{r^{d+1}}, \\
qI(1) &= \int f(\xi)|\xi|^2 e^{-|\xi|^2/2} d\xi - dI(1), \\
(q+d) \int f(\xi)e^{-|\xi|^2/2} d\xi &= \int |\xi|^2 f(\xi)e^{-|\xi|^2/2} d\xi.
\end{aligned}$$

This is the desired (43).

## 4 Analysis of linear map methods

This section contains the calculations behind the results (36) and (39). We estimate the expectations required for  $Q$  (see (3)) using the Laplace asymptotic expansion method, see, e.g., [24]. The calculations are easy to justify if  $F$  has a unique global minimum and  $F \rightarrow \infty$  rapidly enough as  $|x| \rightarrow \infty$ .

### 4.1 Laplace asymptotics, simple linear map

We wish to calculate the expected value of the weight and the expected value of the square of the weight for the linear map method in (8). We use the Taylor expansion of  $F$  in (33) to obtain a Taylor expansion of the weight

$$\begin{aligned}
w(x) &= e^{-\varepsilon^{1/2}C_3(x) - \varepsilon C_4(x) + O(\varepsilon^{3/2})} \\
&= 1 - [\varepsilon^{1/2}C_3(x) + \varepsilon C_4(x)] + \frac{1}{2} [\varepsilon^{1/2}C_3(x)]^2 + O(\varepsilon^{3/2}) \\
&= 1 - \varepsilon^{1/2}C_3(x) + \varepsilon \left[ \frac{1}{2}C_3(x)^2 - C_4(x) \right] + O(\varepsilon^{3/2}). \tag{46}
\end{aligned}$$

Recall that  $C_3(x)$  is an odd function of  $x$  and  $\pi(x)$  is symmetric. Therefore  $E_\pi(C_3) = 0$ , and the variance lemma (41) with  $r = 1/2$  gives

$$Q = \varepsilon \operatorname{var}_\pi(C_3) + O(\varepsilon^{3/2}) = \varepsilon E_\pi(C_3^2) + O(\varepsilon^{3/2}).$$

This is the desired result (36).

## 4.2 Laplace asymptotics, symmetrized linear map

We obtain the Taylor expansion of the weight of the symmetrized linear map method from (15) and from the expansion of the weight of the simple linear map method in (46). We note that the term that is anti-symmetric in  $x$ , which is  $C_3(-x) = -C_3(x)$ , cancels, so that

$$w_s(x) \approx 1 + \varepsilon \left[ \frac{1}{2}C_3^2 - C_4(x) \right] . \quad (47)$$

To apply the variance lemma, we first show that

$$Q = \frac{E_{q_s}(w_s(X)^2)}{E_{q_s}(w_s(X))^2} - 1 = \frac{E_\pi(w_s(\xi)^2)}{E_\pi(w_s(\xi))^2} - 1 . \quad (48)$$

This shows that we can average over  $\xi$  instead of  $X$  when computing the quality measure  $Q$ . To see why, note that (13) implies that for any function  $u$ ,

$$E_{q_s}(u(X)) = E_\pi \left( \frac{2w(\xi)}{w(\xi) + w(-\xi)} u(\xi) \right) . \quad (49)$$

Together with (15), this implies that

$$\begin{aligned} E_{q_s}(w_s(X)) &= E_\pi \left( \frac{2w(\xi)}{w(\xi) + w(-\xi)} w_s(\xi) \right) \\ &= E_\pi \left( \frac{2w(\xi)}{w(\xi) + w(-\xi)} \frac{w(\xi) + w(-\xi)}{2} \right) \\ &= E_\pi(w(\xi)) \end{aligned} \quad (50)$$

The last equality follows from the symmetry of  $\pi$ . Similar algebra and symmetry reasoning leads to

$$\begin{aligned} E_{q_s}(w_s(X)^2) &= E_\pi \left( \frac{2w(\xi)}{w(\xi) + w(-\xi)} \left[ \frac{w(\xi) + w(-\xi)}{2} \right]^2 \right) \\ &= E_\pi \left( w(\xi) \frac{w(\xi) + w(-\xi)}{2} \right) \\ &= E_\pi(w_s(\xi)^2) . \end{aligned} \quad (51)$$

Application of the variance lemma to the above expression, with expectations over  $\xi$ , and using (47), leads to the error term (39).

### 4.3 Evaluating $E(C_3^2)$ with Wick's formula

There is a more explicit expression for  $E_\pi(C_3^2)$  based on Wick's formula [17]. Recall that the distribution of a mean zero multivariate Gaussian is completely determined by its covariance matrix. Therefore, the expected value of a higher order monomial is a function of the covariances. Wick's formula (52) is this function. Of course, the expected value of an odd order monomial is zero; Wick's formula gives the even-order moments.

The general version of Wick's formula is as follows (see, e.g., [17]). Suppose  $X = (X_1, \dots, X_d) \in \mathbb{R}^d$  is a multivariate mean zero Gaussian with covariances  $C_{ij} = E_\pi(X_i X_j)$ . Let  $i_k$ , for  $k = 1, \dots, 2n$ , be a list of indices, with repeats allowed. Let  $M = X_{i_1} \cdots X_{i_{2n}}$  be the corresponding degree  $2n$  monomial. A *pairing* is a partition of  $\{1, \dots, 2n\}$  into  $n$  sets of size 2

$$P = \{\{k_i, l_1\}, \dots, \{k_n, l_n\}\} .$$

A pairing has the property that

$$\{1, \dots, 2n\} = \bigcup_{r=1}^n \{k_r, l_r\} .$$

The set of all pairings is  $\mathcal{P}$ . The number of pairings is

$$|\mathcal{P}| = (2n - 1)(2n - 3) \cdots 3 = (2n - 1)!! .$$

There are no pairings of a set with an odd number of elements. Wick's formula gives the expected value of a monomial of even degree as a sum over all pairings of the indices:

$$E_\pi \left( \prod_{k=1}^{2n} X_{i_k} \right) = \sum_{P \in \mathcal{P}} \prod_{r=1}^n E_\pi(X_{i_{k_r}} X_{i_{l_r}}) = \sum_{P \in \mathcal{P}} \prod_{k=1}^n C_{i_{k_r}, i_{l_r}} . \quad (52)$$

As an example, for  $d = 1$  and  $2n = 6$ ,  $X \sim \mathcal{N}(0, \sigma^2)$ , there are  $5 \cdot 3 = 15$  pairings, so

$$E(X^6) = 15 (\sigma^2)^3 = 15\sigma^6 . \quad (53)$$

To apply Wick's formula to our results, we use the simplified notation  $F_{ijk} = \partial_{x_i} \partial_{x_j} \partial_{x_k} F(0)$ , and write

$$C_3 = \frac{1}{6} \sum_{ijk} F_{ijk} x_i x_j x_k ,$$

and

$$C_3^2 = \frac{1}{36} \sum_{ijklmn} F_{ijk} F_{lmn} x_i x_j x_k x_l x_m x_n .$$

There are two kinds of pairings. One kind pairs one of the indices  $\{i, j, k\}$  with another of the  $\{i, j, k\}$ . This forces one of the  $\{l, m, n\}$  to be paired with another, and the unpaired index from  $\{i, j, k\}$  to be paired with the unpaired index from  $\{l, m, n\}$ . An example of this kind of pairing is

$$P = \{\{i, k\}, \{j, n\}, \{l, m\}\} .$$

There are nine such pairings, since the unpaired index from each triple is arbitrary. The expectations are all equal because  $F_{ijk}$  is a symmetric function of its indices. The other kind of pairing has each of the  $\{i, j, k\}$  paired with one of the  $\{l, m, n\}$ . An example of this kind of pairing is

$$P = \{\{i, m\}, \{j, n\}, \{k, l\}\} .$$

There are six such pairings, since  $i$  is paired with one of the three  $\{l, m, n\}$ , then  $j$  with one of the remaining two, then  $k$  with the last one. The expectations are again equal. Altogether

$$E_\pi(C_3^2) = \frac{1}{36} \sum_{ijklmn} F_{ijk} F_{lmn} (9C_{ij}C_{kl}C_{mn} + 6C_{il}C_{jm}C_{kn}) .$$

This formula simplifies in the special case  $H = I$ , which implies that  $C_{jk} = \delta_{jk}$ . In that case, the  $C_{ij}C_{kl}C_{mn}$  terms vanish unless  $i = j, k = l$  and  $m = n$ . The  $C_{ij}C_{kl}C_{mn}$  terms give

$$\sum_{ikm} F_{iik} F_{mmk} = \|\nabla \triangle F(0)\|_{\ell^2}^2 .$$

(For any tensor  $A$ , we denote the Euclidean 2-norm of all its entries by  $\|A\|_{\ell^2}$  regardless of the rank of  $A$ .) The  $C_{il}C_{jm}C_{kn}$  terms give

$$\sum_{ijk} F_{ijk}^2 = \|D^3 F(0)\|_{\ell^2}^2 .$$

Taken together, these results say that when the Gaussian part of  $p$  is invariant under orthogonal transformations, we have

$$E_\pi(C_3^2) = \frac{1}{4} \|\nabla \triangle F(0)\|_{\ell^2}^2 + \frac{1}{6} \|D^3 F(0)\|_{\ell^2}^2 .$$

The compact expressions on the right represent the two distinct ways a quadratic function of the  $F_{ijk}$  can be rotationally invariant.



## 5 Analysis of random map methods

We analyze the simple and symmetrized random map algorithms in the small noise limit  $\varepsilon \rightarrow 0$ . For the analysis, we use the fact that the random map sampler is affine invariant. This means that if  $M$  is an invertible  $d \times d$  matrix and  $y = Mx$ , then the behavior of the random map sampler is identical when applied to  $F(x)$  or to  $G(x) = F(Mx)$ . Since  $H$  is non-degenerate, it is possible to choose  $M$  so that the Hessian of  $G$  is the identity. Without loss of generality, we put  $H = I$  in our analysis of random map samplers. See [13] for a discussion of the value of affine invariance in practical Monte Carlo.

### 5.1 Simple random map

The powers in  $Q$  for the simple and symmetrized random map methods come easily. The simple method has  $w = 1 + O(\varepsilon^{1/2})$ , which the variance lemma (42) turns into  $Q = O(\varepsilon)$ . The symmetrized method (30) symmetrizes  $w$ , which eliminates the  $O(\varepsilon^{1/2})$  term, leaving  $w_s = O(\varepsilon)$  and  $Q = O(\varepsilon^2)$ . It takes more detailed calculations to find the error constants (37) and (40).

It is clear that with our assumptions  $w$  has an asymptotic expansion in powers of  $\varepsilon^{1/2}$  as required by the variance lemma. For the error constant of the symmetrized method, we need explicit expressions up to  $O(\varepsilon)$ . We calculate the expansions of the quantities that enter into  $w$ , then combine them. We write  $a(\xi, \varepsilon) \approx b(\xi, \varepsilon)$  if  $a$  and  $b$  agree up to order  $\varepsilon$ .

With our normalization  $H = I$ , we obtain from (33)

$$F(x) \approx \frac{1}{2}|x|^2 + \varepsilon^{1/2}C_3(x) + \varepsilon C_4(x). \quad (54)$$

To find an expansion for  $\lambda$ , we substitute the ansatz

$$\lambda(\xi) \approx 1 + \varepsilon^{1/2}\lambda_1(\xi) + \varepsilon\lambda_2(\xi) \quad (55)$$

into (17). We find that

$$\begin{aligned} \frac{1}{2}|\xi|^2 &\approx \frac{1}{2}\lambda^2(\xi)|\xi|^2 + \varepsilon^{1/2}\lambda(\xi)^3C_3(\xi) + \varepsilon\lambda(\xi)^4C_4(\xi), \\ &\approx \frac{1}{2}|\xi|^2 + \varepsilon^{1/2}\lambda_1|\xi|^2 + \varepsilon\left[\frac{1}{2}\lambda_1^2 + \lambda_2\right]|\xi|^2, \\ &\quad + \varepsilon^{1/2}\left[1 + 3\varepsilon^{1/2}\lambda_1\right]C_3(\xi) + \varepsilon C_4(\xi). \end{aligned} \quad (56)$$

Collecting terms of  $O(\varepsilon^{1/2})$  gives

$$0 = \lambda_1(\xi)|\xi|^2 + C_3(\xi) ,$$

which can be rearranged to

$$\lambda_1(\xi) = \frac{-C_3(\xi)}{|\xi|^2} . \quad (57)$$

The  $O(\varepsilon)$  equation is

$$0 = \left[ \frac{1}{2}\lambda_1(\xi)^2 + \lambda_2(\xi) \right] |\xi|^2 + 3\lambda_1(\xi)C_3(\xi) + C_4(\xi) .$$

Solving for  $\lambda_2$  yields

$$\lambda_2(\xi) = \frac{5}{2} \frac{C_3(\xi)^2}{|\xi|^4} - \frac{C_4(\xi)}{|\xi|^2} . \quad (58)$$

We now expand the weights (20). For the denominator, we compute the gradient of  $F$ :

$$\nabla F(\xi) \approx \xi + \varepsilon^{1/2} \nabla C_3(\xi) + \varepsilon \nabla C_4(\xi) .$$

Since  $C_3(\xi)$  is homogeneous of degree 3, we have  $\nabla C_3(\lambda\xi) = \lambda^2 \nabla C_3(\xi)$ , and Euler's identity gives  $\xi^t \nabla C_3(\xi) = 3C_3(\xi)$ . Therefore,

$$\begin{aligned} \xi^t \nabla F(\lambda(\xi)\xi) &\approx \lambda(\xi)|\xi|^2 + \varepsilon^{1/2} \lambda(\xi)^2 \xi^t \nabla C_3(\xi) + \varepsilon \xi^t \nabla C_4(\xi) \\ &\approx |\xi|^2 + \varepsilon^{1/2} [\lambda_1(\xi)|\xi|^2 + 3C_3(\xi)] \\ &\quad + \varepsilon [\lambda_2(\xi)|\xi|^2 + 6\lambda_1(\xi)C_3(\xi) + 4C_4(\xi)] \\ &\approx |\xi|^2 + \varepsilon^{1/2} 2C_3(\xi) + \varepsilon \left[ 3C_4(\xi) - \frac{7}{2} \frac{C_3(\xi)^2}{|\xi|^2} \right] . \end{aligned} \quad (59)$$

For the numerator in (20), use the identity

$$(1 + \alpha)^{d-1} = 1 + (d-1)\alpha + \frac{1}{2}(d-1)(d-2)\alpha^2 + O(\alpha^3) ,$$

to obtain

$$\begin{aligned} \lambda^{d-1} &\approx \left( 1 + \varepsilon^{1/2} \lambda_1 + \varepsilon \lambda_2 \right)^{d-1} \\ &\approx 1 + \varepsilon^{1/2} (d-1) \lambda_1 + \varepsilon \left[ (d-1) \lambda_2 + \frac{1}{2} (d-1)(d-2) \lambda_1^2 \right] \\ &\approx 1 + \varepsilon^{1/2} \frac{(1-d)C_3(\xi)}{|\xi|^2} + \varepsilon (d-1) \left[ \frac{d+3}{2} \frac{C_2(\xi)^2}{|\xi|^4} - \frac{C_4(\xi)}{|\xi|^2} \right] . \end{aligned} \quad (60)$$

We use (60) and (59) to evaluate  $w$  to order  $\varepsilon$ :

$$w(\xi) \approx \frac{\left\{ 1 + \varepsilon^{1/2} \frac{(1-d)C_3(\xi)}{|\xi|^2} + \varepsilon(d-1) \left[ \frac{d+3}{2} \frac{C_2(\xi)^2}{|\xi|^4} - \frac{C_4(\xi)}{|\xi|^2} \right] \right\} |\xi|^2}{|\xi|^2 + \varepsilon^{1/2} 2C_3(\xi) + \varepsilon \left[ 3C_4(\xi) - \frac{7}{2} \frac{C_3(\xi)^2}{|\xi|^2} \right]} .$$

This has the form

$$w \approx \frac{1 + \varepsilon^{1/2}a + \varepsilon b}{1 + \varepsilon^{1/2}c + \varepsilon d} \approx 1 + \varepsilon^{1/2}(a - c) + \varepsilon (c^2 - d - ac + b) , \quad (61)$$

with coefficients

$$\begin{aligned} a &= \frac{(1-d)C_3(\xi)}{|\xi|^2} , \\ b &= (d-1) \left[ \frac{d+3}{2} \frac{C_3(\xi)^2}{|\xi|^4} - \frac{C_4(\xi)}{|\xi|^2} \right] , \\ c &= \frac{2C_3(\xi)}{|\xi|^2} , \\ d &= \frac{3C_4(\xi)}{|\xi|^2} - \frac{7}{2} \frac{C_3(\xi)^2}{|\xi|^4} . \end{aligned}$$

The term of order  $\varepsilon^{1/2}$  in (61) is

$$a - c = -(d+1) \frac{C_3^2(\xi)}{|\xi|^2} . \quad (62)$$

This suffices for the error constant for the simple random map method. The variance lemma formula (42), together with (62) gives

$$Q \approx \varepsilon(d+1)^2 E_\pi \left( \frac{C_3(\xi)^2}{|\xi|^4} \right) . \quad (63)$$

The expected value can be evaluated using the Gaussian integral identity (45). Since  $C_3^2(\xi)$  is degree  $p = 6$ , we have

$$Q \approx \varepsilon \frac{(d+1)^2}{(d+2)(d+4)} E_\pi (C_3(\xi)^2) . \quad (64)$$

This is the desired result (37).

## 5.2 Symmetrized random map

The analysis of the symmetrized random map requires the  $O(\varepsilon)$  term in the  $w$  expansion (61). The result is

$$\frac{(d+2)(d+4)}{2} \frac{C_3(\xi)^2}{|\xi|^4} - (d+2) \frac{C_4(\xi)}{|\xi|^2}.$$

The  $w$  symmetrization formula (30) then gives

$$w_s(\xi) \approx 1 + \varepsilon \left[ \frac{(d+2)(d+4)}{2} \frac{C_3(\xi)^2}{|\xi|^4} - (d+2) \frac{C_4(\xi)}{|\xi|^2} \right].$$

The variance lemma (42) then implies that

$$Q = \varepsilon^2 \text{var}_\pi \left( \frac{(d+2)(d+4)}{2} \frac{C_3(\xi)^2}{|\xi|^4} - (d+2) \frac{C_4(\xi)}{|\xi|^2} \right).$$

We now expand the above, and rearrange the terms for direct comparison with the result (39) for the symmetrized linear map:

$$\begin{aligned} Q &= \underbrace{\varepsilon^2 \text{var}_\pi \left( \frac{(d+2)(d+4)}{2} \cdot \frac{C_3(\xi)^2}{|\xi|^4} \right)}_{\text{I}} \\ &\quad - \underbrace{\varepsilon^2 \text{cov}_\pi \left( \frac{(d+2)(d+4)}{2} \cdot \frac{C_3(\xi)^2}{|\xi|^4}, (d+2) \frac{C_4(\xi)}{|\xi|^2} \right)}_{\text{II}} \\ &\quad + \underbrace{\varepsilon^2 \text{var}_\pi \left( (d+2) \frac{C_4(\xi)}{|\xi|^2} \right)}_{\text{III}}. \end{aligned} \tag{65}$$

Consider term I. We have

$$\begin{aligned} \text{I} &= \text{var}_\pi \left( \frac{(d+2)(d+4)}{2} \frac{C_3(\xi)^2}{|\xi|^4} \right) \\ &= \frac{(d+2)^2(d+4)^2}{4} E_\pi \left( \frac{C_3(\xi)^4}{|\xi|^8} \right) - \left[ \frac{(d+2)(d+4)}{2} E_\pi \left( \frac{C_3(\xi)^2}{|\xi|^4} \right) \right]^2. \end{aligned} \tag{66}$$

Using (45) and a direct generalization of it, we get

$$E_\pi\left(\frac{C_3(\xi)^2}{|\xi|^4}\right) = \frac{E_\pi\left(C_3(\xi)^2\right)}{(d+2)(d+4)}, \quad (67)$$

$$E_\pi\left(\frac{C_3(\xi)^4}{|\xi|^8}\right) = \frac{E_\pi\left(C_3(\xi)^4\right)}{(d+4)(d+6)(d+8)(d+10)}. \quad (68)$$

Thus,

$$\begin{aligned} \text{I} &= \frac{(d+2)^2(d+4)^2 E_\pi\left(C_3(\xi)^4\right)}{4(d+4)(d+6)(d+8)(d+10)} - \left[\frac{1}{2} E_\pi\left(C_3(\xi)^2\right)\right]^2 \\ &= \frac{1}{4} \text{var}_\pi(C_3(\xi)^2) - \frac{1}{4} \cdot \left\{1 - \frac{(d+2)^2(d+4)^2}{(d+4)(d+6)(d+8)(d+10)}\right\} E_\pi(C_3(\xi)^4). \end{aligned} \quad (69)$$

Similarly, we have

$$\text{II} = \frac{1}{2} \text{cov}_\pi\left(C_3(\xi)^2, C_4(\xi)\right) - \frac{1}{2} \cdot \left\{1 - \frac{(d+2)^2(d+4)}{(d+4)(d+6)(d+8)}\right\} E_\pi(C_3(\xi)^2 C_4(\xi)), \quad (70)$$

$$\text{III} = \text{var}_\pi(C_4(\xi)) - \left\{1 - \frac{(d+2)^2}{(d+4)(d+6)}\right\} E_\pi(C_4(\xi)^2). \quad (71)$$

When  $d$  is sufficiently large, we can rewrite the expressions for I-III in a more concise way:

$$\text{I} = \frac{1}{4} \text{var}_\pi(C_3(\xi)^2) - \left(\frac{4}{d} + O(1/d^2)\right) E_\pi(C_3(\xi)^4), \quad (72)$$

$$\text{II} = \frac{1}{2} \text{cov}_\pi\left(C_3(\xi)^2, C_4(\xi)\right) - \left(\frac{5}{d} + O(1/d^2)\right) E_\pi(C_3(\xi)^2 C_4(\xi)), \quad (73)$$

$$\text{III} = \text{var}_\pi(C_4(\xi)) - \left(\frac{6}{d} + O(1/d^2)\right) E_\pi(C_4(\xi)^2). \quad (74)$$

This leads to

$$Q = \varepsilon^2 \left[ \text{var}_\pi\left(\frac{1}{2} C_3(\xi)^2 - C_4(\xi)\right) + c_d \cdot K \right] + O(\varepsilon^3), \quad (75)$$

where  $c_d = O(1/d)$  and  $K$  is a combination of moments of  $C_3$  and  $C_4$ . This verifies the stated result (40). We see that for  $d \gg 1$ , the variance of the symmetrized random map method approaches that of the symmetrized linear map. The above also shows that in low dimensions, the variance of the symmetrized random map may be smaller than that of the symmetrized linear map, though exactly how much depends on the degree of correlation between  $C_3(\xi)^2$  and  $C_4(\xi)$ .

The error term of the symmetrized methods in (39) and (40) can in principle also be evaluated using Wick's formula, however the calculations are much more involved. We illustrate how to use Wick's formula for the error terms of the symmetrized methods with an example.

## 6 Computational experiments

We present computational experiments that confirm the theoretical error analysis above for small  $\varepsilon$  and suggest what may happen when  $\varepsilon$  is not so small. We use two test problems. One is a nonlinear random walk whose dimension is arbitrary. This allows us to see how the samplers' performance depends on the dimension. We see that the samplers perform worse in higher dimension, but they are still quite useful in dimensions of practical interest. In the other example we apply the algorithms to a data assimilation problem with the "Lorenz '63" model [20]. The goal is to sample the posterior distribution of the initial conditions in the presence of noisy observations of the state at later time.

### 6.1 Non-linear random walk

Consider a non-Gaussian random walk tied at the start and free at the end. The random variable is  $X = (X_1, \dots, X_N)$ , with  $X_0 \equiv 0$  implicitly. The Gaussian random walk potential is

$$x^t H x = \sum_{k=0}^{N-1} (x_{k+1} - x_k)^2 \quad (76)$$

where  $x_0 = 0$ . We make the walk non-Gaussian by adding cubic and quartic terms to the potential energy. The nonlinear parts are discretizations of

a nonlinear energy functional

$$C_3(x) = \alpha \sum_{k=0}^{N-1} (x_{k+1} - x_k)^3 , \quad (77)$$

and

$$C_4(x) = \beta \sum_{k=0}^{N-1} (x_{k+1} - x_k)^4 . \quad (78)$$

The coefficients  $\alpha$  and  $\beta$  would be called “coupling constants” in field theory, and both are set to 1 in our numerical experiments below. In the Gaussian measure determined by (76), the increments  $(X_{k+1} - X_k)$  are independent standard normal random variables. We can use this to calculate

$$E_\pi(C_3(X)^2) = \alpha^2 \sum_{jk} E_\pi((X_{j+1} - X_j)^3 (X_{k+1} - X_k)^3) . \quad (79)$$

The terms on the right hand side with  $j \neq k$  vanish because the increments are independent. The terms with  $j = k$  satisfy, using Wick’s formula (53),

$$E_\pi((X_{k+1} - X_k)^6) = E_{\mathcal{N}(0,1)}(Z^6) = 15 , \quad (80)$$

so

$$E_\pi(C_3(x)^2) = 15\alpha^2 N . \quad (81)$$

Thus, the simple linear method for this problem has the quality measure, see (36),

$$Q \approx \varepsilon 15\alpha^2 N . \quad (82)$$

The simple random map quality measure (37) is slightly less:

$$Q \approx \varepsilon 15\alpha^2 \frac{N(N+1)^2}{(N+2)(N+4)} . \quad (83)$$

It is tedious but straightforward to calculate the error constant for the symmetrized methods. We need

$$\text{var}_\pi(C_4 - \frac{1}{2}C_3^2) = E_\pi([C_4 - \frac{1}{2}C_3^2]^2) - (E_\pi(C_4 - \frac{1}{2}C_3^2))^2 .$$

The first part is

$$E_\pi([C_4 - \frac{1}{2}C_3^2]^2) = E_\pi(C_4^2) - E_\pi(C_4 C_3^2) + \frac{1}{4}E_\pi(C_3^4) .$$

We evaluate these three using Wick identities, first

$$\begin{aligned}
E_\pi (C_4^2) &= \beta^2 \sum_{jk} E_{\mathcal{N}(0,1)} ((X_{j+1} - X_j)^4 (X_{k+1} - X_k)^4) \\
&= \beta^2 \left[ \sum_{j \neq k} E ((X_{j+1} - X_j)^4 (X_{k+1} - X_k)^4) \right. \\
&\quad \left. + \sum_{j=k} E ((X_{j+1} - X_j)^4 (X_{k+1} - X_k)^4) \right] \\
&= \beta^2 \left[ (N^2 - N) E ((X_2 - X_1)^4 (X_3 - X_2)^4) \right. \\
&\quad \left. + N E ((X_2 - X_1)^8) \right] \\
&= 3 \cdot 3 \beta^2 N^2 + O(N) .
\end{aligned}$$

We write numbers in factored form, as in  $3 \cdot 3$  instead of 9, for clarity. The second term is

$$\begin{aligned}
E_\pi (C_4 C_3^2) &= \alpha^2 \beta \sum_{jkl} E_{\mathcal{N}(0,1)} ((X_{j+1} - X_j)^4 (X_{k+1} - X_k)^3 (X_{l+1} - X_l)^3) \\
&= \alpha^2 \beta \left[ \sum_{j \neq (k=l)} E ((X_{j+1} - X_j)^4 (X_{k+1} - X_k)^3 (X_{l+1} - X_l)^3) \right. \\
&\quad \left. + \sum_{j=k=l} E ((X_{j+1} - X_j)^4 (X_{k+1} - X_k)^3 (X_{l+1} - X_l)^3) \right] \\
&= \alpha^2 \beta \left[ (N^2 - N) E ((X_2 - X_1)^4 (X_3 - X_2)^6) \right. \\
&\quad \left. + N E ((X_2 - X_1)^{10}) \right] \\
&= 3 \cdot 5 \cdot 3 \alpha^2 \beta N^2 + O(N) .
\end{aligned}$$

The factor of 3 in the third term is for the three possibilities  $(j = k) \neq (l =$



$m)$ , and  $(j = l) \neq (k = m)$ , and  $(j = m) \neq (k = l)$ :

$$\begin{aligned}
E_\pi(C_3^4) &= \alpha^4 \sum_{jklm} E((X_{j+1} - X_j)^3 (X_{k+1} - X_k)^3 (X_{l+1} - X_l)^3 (X_{m+1} - X_m)^3) \\
&= \alpha^4 \left[ 3 \sum_{(j=k) \neq (l=m)} E((X_{j+1} - X_j)^3 (X_{k+1} - X_k)^3 (X_{l+1} - X_l)^3 (X_{m+1} - X_m)^3) \right. \\
&\quad \left. + \sum_{j=k=l=m} E((X_{j+1} - X_j)^3 (X_{k+1} - X_k)^3 (X_{l+1} - X_l)^3 (X_{m+1} - X_m)^3) \right] \\
&= \alpha^4 \left[ 3(N^2 - N) E((X_2 - X_1)^6 (X_3 - X_2)^6) \right. \\
&\quad \left. + N E((X_2 - X_1)^{12}) \right] \\
&= 3 \cdot (5 \cdot 3)^2 \alpha^4 N^2 + O(N) .
\end{aligned} \tag{84}$$

Adding these gives

$$E_\pi([C_4 - \frac{1}{2}C_3^2]^2) = N^2 (\beta^2 \cdot 3^2 - \alpha^2 \beta \cdot 3 \cdot 5 \cdot 3 + \frac{1}{4}\alpha^4 \cdot 3 \cdot (5 \cdot 3)^2) + O(N) .$$

A simpler calculation shows that  $E_\pi(C_4 - \frac{1}{2}C_3^2) = O(N)$ . Subtracting the terms finally gives

$$\text{var}_\pi(C_4 - \frac{1}{2}C_3^2) = \frac{1}{4}\alpha^4 N^2 \cdot 2 \cdot (5 \cdot 3)^2 + O(N) = \frac{225\alpha^4 N^2}{2} + O(N) .$$

It is now clear that the simple methods have error coefficients proportional to  $\varepsilon N$ , and the symmetrized methods have error coefficients proportional to  $(\varepsilon N)^2$ .

We perform numerical experiments and vary  $N$  and  $\varepsilon$ . In these experiments, we approximate the expected values in the quality measure  $Q$  by averages over  $10^4$  samples. We protect the computations against over- and underflow as follows. Instead of saving the weight, we save the logarithm of the weight of each sample. This is straightforward for the linear map. For the symmetrized linear map, we use

$$\begin{aligned}
w_{slm}(x) &\propto w_{lm}(x) + w_{lm}(-x) \\
&= w_{lm}(x) \left( 1 + \frac{w_{lm}(-x)}{w_{lm}(x)} \right) ,
\end{aligned} \tag{85}$$

where  $w_{lm}$  is the weight of the simple linear map and  $w_{slm}$  that of the symmetrized linear map. We then compute

$$\begin{aligned}\log w_{slm}(x) &= \log(w_{lm}(x)) + \log\left(1 + \frac{w_{lm}(-x)}{w_{lm}(x)}\right), \\ &= \log(w_{lm}(x)) + \log(1 + \exp(F(x) - F(-x))).\end{aligned}\quad (86)$$

For the random map we save the log of the weight

$$\log w_{rm}(x) = (d-1)\log(|\lambda(x)|) + \log(\xi^t H \xi) - \log(\xi^t \nabla_x F(\lambda(x)\xi)). \quad (87)$$

For the symmetrized random map, the log of the weight is

$$\begin{aligned}\log w_{srm}(x) &= \log(w_{rm}(x)), \\ &+ \log\left(1 + \left(\frac{\lambda(-x)}{\lambda(x)}\right)^{d-1} \frac{\xi^t \nabla_x F(-\lambda(-x)\xi)}{\xi^t \nabla_x F(\lambda(x)\xi)}\right).\end{aligned}\quad (88)$$

Once we have computed the logarithms of the weights for each sample, we subtract the maximum value of the logarithms of the weights, then exponentiate, then normalize.

The left panel of Figure 1 shows  $Q$  as a function of  $\varepsilon$  for  $N = 2$ , and the right panel for  $N = 200$ . The dots, circles, squares and diamonds are values of  $Q$  computed from the samples, the lines have slope one or two, and are there to illustrate the “order” of the method. Specifically, the turquoise line is as in (82), the red line as in (83), and the purple line on the left is as in (84). The numerical results confirm our asymptotic expansions for sufficiently small  $\varepsilon$ . We have made similar observations for other values of  $N \leq 1000$ . Specifically, for  $N = 2$ , we observe that the numerical results agree with the predicted values for relatively large  $\varepsilon$  (up to  $\varepsilon \approx 0.01$ ). For  $\varepsilon \geq 10^{-3}$ , the linear map method, the random map method, and the symmetrized linear map method are similarly good (as measured by  $Q$ ). All four methods are doing equally well when  $\varepsilon$  becomes even larger. Moreover, all four methods can be useful in this problem, in the sense that  $Q$  is “not too large,” even when  $\varepsilon$  is close to 1.

We observe in the numerical experiments with  $N = 200$  that the random map loses its advantage over the linear map when  $N$  becomes large. This is true for the simple and symmetrized versions of these methods. We observe that the results of our experiments agree with our predictions for

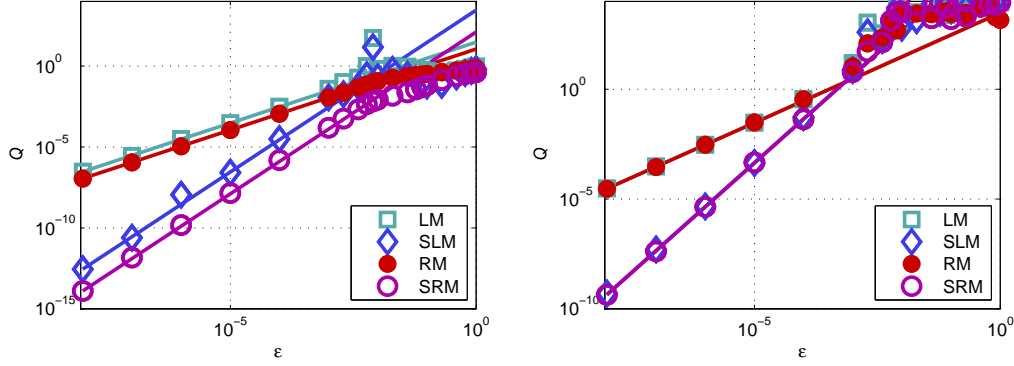


Figure 1: Sampling nonlinear random walks. Left:  $N = 2$ . Right:  $N = 200$ . Turquoise squares: linear map (LM). Blue diamonds: symmetrized linear map (SLM). Red dots: random map. Purple circles: symmetrized random map. The turquoise and red lines have slope one, the purple and blue lines have slope two. The turquoise line is as in (82), the red line as in (83), and the purple line on the left is as in (84).

$\varepsilon \leq 10^{-3}$ . For larger  $\varepsilon$ , all methods perform poorly and yield a large  $Q \gg 1$  for  $\varepsilon \geq 10^{-3}$ .

Figure 2 illustrates the scaling of  $Q$  with  $N$ , as computed by Wick's formula. Shown is  $Q$  as a function of  $N$  for the various methods. As predicted by the theory, we observe that the symmetrized methods have leading error terms proportional to  $(\varepsilon N)^2$ , and that the simple, unsymmetrized methods have error terms proportional to  $\varepsilon N$ .

## 6.2 Lorenz '63

We consider estimating the initial conditions of the Lorenz '63 [20] equations

$$\frac{dx}{dt} = \sigma(y - x), \quad \frac{dy}{dt} = x(\rho - z) - y, \quad \frac{dz}{dt} = xy - \beta z, \quad (89)$$

where  $\sigma = 10$ ,  $\beta = 8/3$  and  $\rho = 28$ , from noisy measurements of  $x$ ,  $y$  and  $z$  at time  $T$ :

$$d = (x(T), y(T), z(T))^t + v. \quad (90)$$

The Gaussian random variable  $v \sim \mathcal{N}(0, \varepsilon I_3)$  models measurement noise. The above ordinary differential equations (ODE) are solved with the Mat-

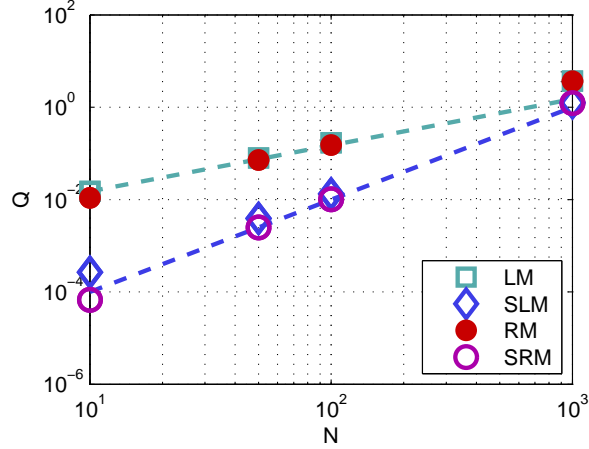


Figure 2: Scaling of  $Q$  with  $N$ . Turquoise squares: linear map (LM). Blue diamonds: symmetrized linear map (SLM). Red dots: random map. Purple circles: symmetrized random map. The turquoise line has slope one, the blue line has slope two.

lab routine `ode45`. The prior for the initial conditions is Gaussian with mean

$$\mu_0 = (3.6314, 6.6136, 10.6044)^t, \quad (91)$$

and covariance  $P_0 = \varepsilon I_3$ . The conditional random variable  $x_0|d$  thus has the pdf  $p(x_0|d) = \exp(-F(x_0)/\varepsilon)$ , where

$$F(x_0) = \frac{1}{2} \left( (d - h(x_0))^t (d - h(x_0)) + (\mu_0 - x_0)^t (\mu_0 - x_0) \right), \quad (92)$$

so that this problem corresponds to a “small noise” situation. Here  $x_0$  is shorthand notation for the vector  $(x(0), y(0), z(0))^t$ , and  $h(x_0)$  is the `ode45` solution of the ODEs at time  $T$ . The initial conditions we use to generate the synthetic data for our numerical experiments is

$$x_{0,\text{true}} = \mu_0 + 0.5 (\sqrt{\varepsilon}, -\sqrt{\varepsilon}, \sqrt{\varepsilon})^t$$

We generate samples of  $x_0|d$  using the linear and random map methods described above, and vary  $\varepsilon$  and  $T$ . The minimization required by the sampling schemes is done with a quasi-Newton method where all derivatives are approximated with finite differences. Similarly, we approximate the Hessian at the minimum via finite differences.

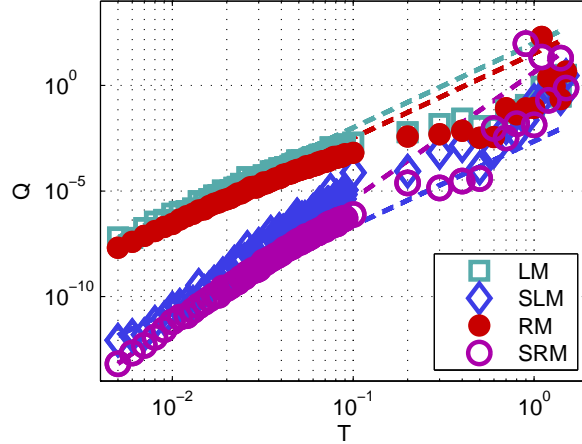


Figure 3: Estimating initial conditions of the Lorenz '63 equations. The parameter  $\varepsilon = 1$  is constant and the time  $T$  at which data are collected is varied. Turquoise squares: linear map (LM). Blue diamonds: symmetrized linear map (SLM). Red dots: random map. Purple circles: symmetrized random map. The turquoise and red lines have slope four, the blue and purple lines have slope six.

We first fix  $\varepsilon = 1$  and vary  $T$ , i.e. the time when data are collected. As  $T$  becomes larger, the problem becomes more and more difficult and multiple modes can appear [21, 22]. Figure 3 shows  $Q$  as a function of  $T$ . We observe that the symmetrized methods perform better than the simple versions, and give a significantly smaller  $Q$ -value. For small  $T$ , the computed values of  $Q$  follow a straight line with slope 4 for the random and linear maps, and slope 6 for the symmetrized methods. For  $T \approx 1$ , all four methods perform similarly well (the symmetrization seems to lose its advantages) and for  $T > 1$ , the methods perform poorly. This is perhaps because the pdf we attempt to sample becomes multi-modal and, therefore, is no longer star-shaped. However, we made no adjustments to address multi-modal target densities.

Next, we fix  $T = 0.05$  and vary  $\varepsilon$ . In this case, the pdf has the functional form we analyze, and the scenario is analogous to the “small noise accurate data” regime analyzed in the context of particle filtering in [28]. Figure 4 shows  $Q$  as a function of  $\varepsilon$ . As in the previous example, we find

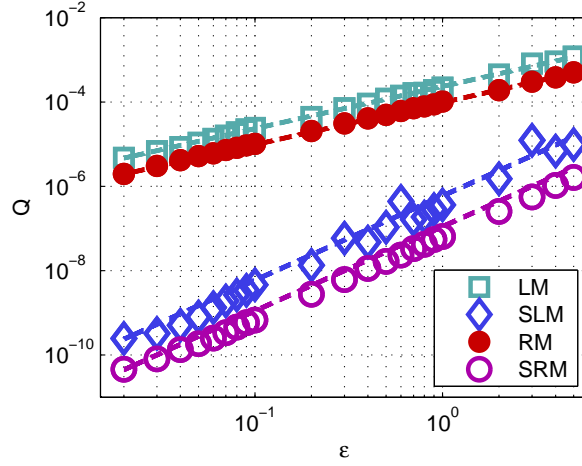


Figure 4: Estimating initial conditions of the Lorenz '63 equations. The data are collected at time  $T = 0.05$  and the parameter  $\varepsilon$  is varied. Turquoise squares: linear map (LM). Blue diamonds: symmetrized linear map (SLM). Red dots: random map. Purple circles: symmetrized random map. The turquoise and red lines have slope one, the blue and purple lines have slope two.

that our numerical experiments confirm the predicted behavior, even if  $\varepsilon$  is relatively large.

## 7 Conclusion and discussion

We have performed a small-noise analysis of two implicit sampling methods, the linear and random map methods. The analysis shows that the random map method outperforms the linear map method in the small noise regime, but this advantage becomes insignificant in high dimensions. The simplicity and relative speed of the linear map method thus makes it more attractive in the limit of small noise. The analysis further suggests that both methods may be improved by a symmetrization procedure analogous to antithetic variates. We illustrate the theory with numerical examples which also suggest that the symmetrized algorithms may outperform the simple, unsymmetrized algorithms even when the noise

is not so small.

We wish to emphasize two points that are important in practice. The first concerns weighted direct samplers as used in particle filtering. Some methods proposed for practical applications do not have  $Q \rightarrow 0$ , and may even have  $Q \rightarrow \infty$  as  $\varepsilon \rightarrow 0$ . For example, the “vanilla” bootstrap particle filter [14], which proposes samples from a proposal distribution that does not take into account the most recent observation, has  $Q \rightarrow \infty$  [27]. The present samplers all make proposals centered about the MAP (maximum a-posteriori) point, which takes into account the most recent observation. There is much discussion in the literature of the advantages of doing this [7, 11, 18, 25, 29].

Second, we wish to address the computational cost of the algorithms we analyze and propose. In practice, the cost is roughly proportional to the number of evaluations of  $F$  and its derivatives. Even our Lorenz ‘63 example requires an ODE solve to evaluate  $F$ . The rest of the algorithm is cheap by comparison.

All of our methods start with computing  $x_* = \operatorname{argmin} F(x)$ . This requires a number of evaluations of  $F$  and possibly its derivatives (for numerical optimization). We also need the Hessian of  $F$ , either by formulas, adjoints, or by finite differences. The simple linear map method requires one more evaluation of  $F(X)$  per sample. The symmetrized linear map method requires two  $F$  evaluations. In particle filter applications, we may want just one sample. In that case, the optimization is more expensive than sampling. Other applications may require many samples, in which case the cost is roughly the number of samples times the cost for one or two  $F$  evaluations.

The simple random map must solve (16) once for each sample. This is one equation in the single unknown,  $\lambda$ . It is normally solved with just a few  $F$  evaluations. As with the linear map methods, generating one sample using the symmetrized method requires roughly the work of two samples from the simple version, though by exploiting the ansatz (55) for  $\lambda$ , one can obtain a good initial guess for finding  $\lambda(-\xi)$  based on  $\lambda(\xi)$ . This may speed up the symmetrized method.

## Acknowledgments

We gratefully acknowledge the influence of Alexandre Chorin on our work, and very helpful conversations with him on this material. Jonathan Goodman and Kevin Lin thank the Lawrence Berkeley National Laboratory for facilitating our collaboration on this project.

The work of Jonathan Goodman was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics Program under contract number DE-AC02005CH11231 under a subcontract from Lawrence Berkeley National Laboratory to New York University. The work of Kevin Lin was supported in part by the National Science Foundation under grants DMS-1217065 and DMS-1418775. The work of Matthias Morzfeld was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics Program under contract number DE-AC02005CH11231, and by the National Science Foundation under grant DMS-1217065.

## References

- [1] S. AN AND F. SCHORFHEIDE, *Bayesian analysis of DSGE models*, *Econometric Reviews*, 26 (2007), pp. 113–172.
- [2] M. ARULAMPALAM, S. MASKELL, N. GORDON, AND T. CLAPP, *A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking*, *IEEE Transactions on Signal Processing*, 50 (2002), pp. 174–188.
- [3] E. ATKINS, M. MORZFELD, AND A. CHORIN, *Implicit particle methods and their connection with variational data assimilation*, *Monthly Weather Review*, 141 (2013), pp. 1786–1803.
- [4] N. BERGMAN, ed., *Recursive Bayesian estimation: Navigation and tracking applications*, Ph.D Dissertation, Linkoping University, Linkoping, Sweden, 1999.
- [5] M. BOCQUET, C. PIRES, AND L. WU, *Beyond Gaussian statistical modeling in geophysical data assimilation*, *Monthly Weather Review*, 138 (2010), pp. 2997–3023.



- [6] A. CHORIN AND O. HALD, *Stochastic Tools in Mathematics and Science*, Springer, third ed., 2013.
- [7] A. CHORIN AND M. MORZFELD, *Conditions for successful data assimilation*, *Journal of Geophysical Research*, 118 (2013), pp. 11,522–11,533.
- [8] A. CHORIN, M. MORZFELD, AND X. TU, *Implicit particle filters for data assimilation*, *Communications in Applied Mathematics and Computational Science*, 5 (2010), pp. 221–240.
- [9] A. CHORIN AND X. TU, *Implicit sampling for particle filters*, *Proceedings of the National Academy of Sciences*, 106 (2009), pp. 17249–17254.
- [10] A. DOUCET, N. DE FREITAS, AND N. GORDON, eds., *Sequential Monte Carlo Methods in Practice*, Springer, 2001.
- [11] A. DOUCET, S. GODSILL, AND C. ANDRIEU, *On sequential Monte Carlo sampling methods for Bayesian filtering*, *Statistics and Computing*, 10 (2000), pp. 197–208.
- [12] A. FOURNIER, G. HULOT, D. JAULT, W. KUANG, W. TANGBORN, N. GILLET, E. CANET, J. AUBERT, AND F. LHUILLIER, *And introduction to data assimilation and*, *Space Science Review*, 137 (2009), pp. 247–291.
- [13] J. GOODMAN AND J. WEARE, *Ensemble samplers with affine invariance*, *Communications in Applied Mathematics and Computational Science*, 5 (2010), pp. 65–80.
- [14] N. GORDON, D. SALMOND, AND A. SMITH, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, *Radar and Signal Processing*, *IEEE Proceedings F*, 140 (1993), pp. 107–113.
- [15] J. HAMMERSLEY AND D. HANDSCOMB, *Monte Carlo Methods*, Chapman & Hall, 1964.
- [16] M. KALOS AND P. WHITLOCK, *Monte Carlo Methods*, vol. 1, John Wiley & Sons, 1 ed., 1986.
- [17] L. KOOPMANS, *The spectral analysis of time series*, Academic Press, 1974.

- [18] J. LIU AND R. CHEN, *Blind deconvolution via sequential imputations*, Journal of the American Statistical Association, 90 (1995), pp. 567–576.
- [19] ———, *Sequential Monte Carlo methods for dynamical systems*, Journal of the American Statistical Association, 93 (1998), pp. 1032–1044.
- [20] E. LORENZ, *Deterministic nonperiodic flow*, Journal of the Atmospheric Sciences, 20 (1963), pp. 130–141.
- [21] R. MILLER, J. CARTER, AND S. BLUE, *Data assimilation into nonlinear stochastic models*, Tellus, 51 (1999), pp. 167–194.
- [22] R. MILLER, M. GHIL, AND F. GAUTHIEZ, *Advanced data assimilation in strongly nonlinear dynamical systems*, Journal of Atmospheric Science, 51 (1994), pp. 1037–1056.
- [23] M. MORZFELD, X. TU, E. ATKINS, AND A. CHORIN, *A random map implementation of implicit filters*, Journal of Computational Physics, 231 (2012), pp. 2049–2066.
- [24] J. MURRAY, *Asymptotic Analysis*, Springer Verlag, 1992.
- [25] C. SNYDER, T. BENGTSSON, P. BICKEL, AND J. ANDERSON, *Obstacles to high-dimensional particle filtering*, Monthly Weather Review, 136 (2008), pp. 4629–4640.
- [26] P. VAN LEEUWEN, *Particle filtering in geophysical systems*, Monthly Weather Review, 137 (2009), pp. 4089–4114.
- [27] E. VANDEN-EIJNDEN AND J. WEARE, *Rare event simulation and small noise diffusions*, Communications on Pure and Applied Mathematics, 65 (2012), pp. 1770–1803.
- [28] ———, *Data assimilation in the low noise, accurate observation regime with application to the kuroshio current*, Monthly Weather Review, (2013).
- [29] V. ZARITSKII AND L. SHIMELEVICH, *Monte Carlo technique in problems of optimal data processing*, Automation and Remote Control, 12 (1975), pp. 95 – 103.